nature physics

ARTICLES

PUBLISHED ONLINE: 10 AUGUST 2015 | DOI: 10.1038/NPHYS3422

Spectrum of controlling and observing complex networks

Gang Yan[1†], Georgios Tsekenis[1†], Baruch Barzel[2], Jean-Jacques Slotine[3,4], Yang-Yu Liu[5,6] and Albert-László Barabási[1,6,7,8*]

Recent studies have made important advances in identifying sensor or driver nodes, through which we can observe or control a complex system. But the observational uncertainty induced by measurement noise and the energy required for control continue to be significant challenges in practical applications. Here we show that the variability of control energy and observational uncertainty for different directions of the state space depend strongly on the number of driver nodes. In particular, we find that if all nodes are directly driven, control is energetically feasible, as the maximum energy increases sublinearly with the system size. If, however, we aim to control a system through a single node, control in some directions is energetically prohibitive, increasing exponentially with the system size. For the cases in between, the maximum energy decays exponentially when the number of driver nodes increases. We validate our findings in several model and real networks, arriving at a series of fundamental laws to describe the control energy that together deepen our understanding of complex systems.

Many natural and man-made systems can be represented as networks[1–3], where nodes are the system's components and links describe the interactions between them. Thanks to these interactions, perturbations of one node can alter the states of the other nodes[4–6]. This property has been exploited to control a network—that is, to move it from an initial state to a desired final state[7–9]—by manipulating the state variables of only a subset of its nodes[10,11]. Such control processes[10–26] play an important role in the regulation of protein expression[27], the coordination of moving robots[28], and the inhibition of undesirable social contagions[29]. At the same time the interdependence between nodes means that the states of a small number of sensor nodes contain sufficient information about the rest of the network, so that we can reconstruct the system's full internal state by accessing only a few outputs[30]. This can be utilized for biomarker design in cellular networks, or to monitor in real time the state and functionality of infrastructural[31] and social–ecological[32] systems for early warning of failures or disasters[33].

Although recent advances in driver and sensor node identification constitute unavoidable steps towards controlling and observing real networks, in practice we continue to face significant challenges: the control of a large network may require a vast amount of energy[16–18], and measurement noise[34] causes uncertainties in the observation process. To quantify these issues we formalize the dynamics of a controlled network with N nodes and N_D external control inputs as[7–10]

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) \tag{1}$$

where the vector $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T$ describes the states of the N nodes at time t and $x_i(t)$ can represent the concentration of a metabolite in a metabolic network[35], the geometric state of a chromosome in a chromosomal interaction network[14], or the belief of an individual in opinion dynamics[29,36]. The vector $\mathbf{u}(t) = [u_1(t), u_2(t), \ldots, u_{N_D}(t)]^T$ represents the external control inputs, and B is the input matrix, with $B_{ij} = 1$ if control input $u_j(t)$ is imposed on node i. The adjacency matrix A captures the interactions between the nodes, including the possibility of self-loops A_{ii} representing the self-regulation of node i.

Control energy

The system (1) can be driven from an initial state $\mathbf{x}_o$ to any desired final state $\mathbf{x}_d$ within the time $t \in [0, \tau]$ using an infinite number of possible control inputs $\mathbf{u}(t)$. The optimal input vector aims to minimize the control energy[7] $\int_0^\tau \|\mathbf{u}(t)\|^2 dt$, which captures the energy of electronic and mechanical systems or the amount of effort required to control biological and social systems. If at $t = 0$ the system is in state $\mathbf{x}_o = 0$, the minimum energy required to move the system to point $\mathbf{x}_d$ in the state space can be shown to be[7,16–18]

$$\mathcal{E}(\tau) = \mathbf{x}_d^T G_c^{-1}(\tau)\mathbf{x}_d \tag{2}$$

where $G_c(\tau) = \int_0^\tau e^{At} BB^T e^{A^T t} dt$ is the symmetric controllability Gramian. When the system is controllable all eigenvalues of $G_c(\tau)$ are positive. Equation (2) indicates that for a network A and an input matrix B the control energy $\mathcal{E}(\tau)$ also depends on the desired state $\mathbf{x}_d$. Consequently, driving a network to various directions in the state space requires different amounts of energy. For example, to move the weighted network of Fig. 1a to the three different final states $\mathbf{x}_d$ with $\|\mathbf{x}_d\| = 1$, we inject the optimal signals $\mathbf{u}(t)$ shown in Fig. 1b onto node 1, steering the system along the trajectories shown in

[1]Center for Complex Network Research and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA. [2]Department of Mathematics, Bar-Ilan University, Ramat-Gan 52900, Israel. [3]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA. [7]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [8]Center for Network Science, Central European University, H-1051 Budapest, Hungary. †These authors contributed equally to this work. *e-mail: alb@neu.edu

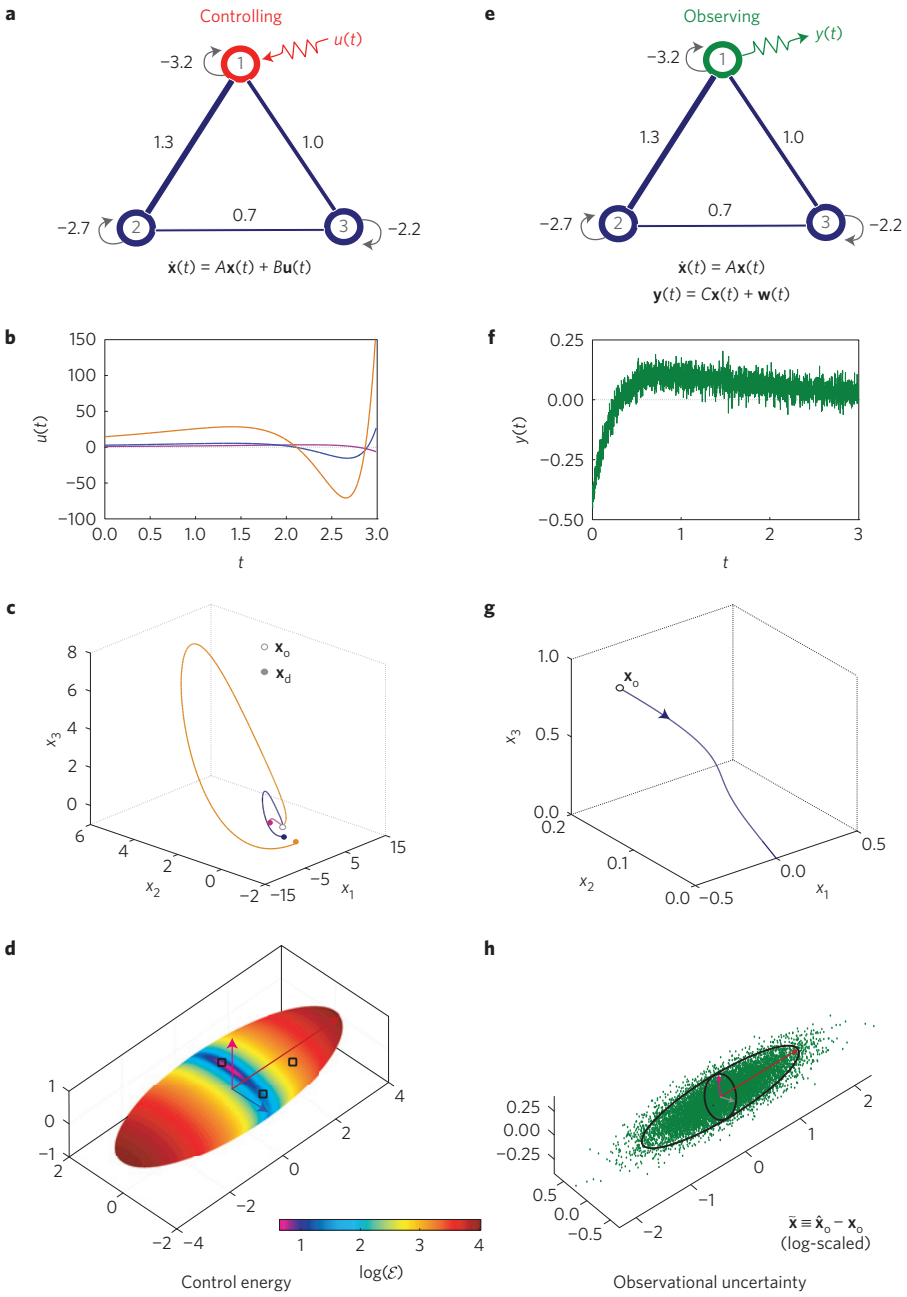

Figure 1 | Controlling and observing a network. a, The control of a three-node weighted network with one external signal $u(t)$ that is injected to the red driver node. Hence the input matrix is $B=[1,0,0]^{\mathrm{T}}$. The nodes have negative self-loops that make all eigenvalues of the adjacency matrix A negative. **b**, Optimal control signals that minimize the energies required to move the network from the initial state $\mathbf{x}_{\mathrm{o}}=[0,0,0]^{\mathrm{T}}$ to three different desired states $\mathbf{x}_{\mathrm{d}}$ with $\|\mathbf{x}_{\mathrm{d}}\|=1$ in the given time interval $t\in[0,3]$. **c**, Trajectories of the network state $\mathbf{x}(t)$ driven respectively by the control signals in **b**. **d**, Control energy surface, showing the amount of energy required to move the network by one unit distance (that is, $\|\mathbf{x}_{\mathrm{d}}\|=1$) in different directions. The surface is an ellipsoid spanned by the eigen-energies for the controllability Gramian's three eigen-directions (arrows). The squares correspond to the three final points used in **b** and **c**. **e**, Observing the network with one output $y(t)$. Node 1 is selected as the sensor (green), thus the output matrix is $C=[1,0,0]$. The measurement noise $\mathbf{w}(t)$ is assumed as Gaussian white noise with zero mean and variance one. **f**, A typical output $y(t)$ that is used to approximate the initial state $\mathbf{x}_{\mathrm{o}}$. **g**, A typical trajectory of the system state $\mathbf{x}(t)$. **h**, Estimation error $\tilde{\mathbf{x}}=\hat{\mathbf{x}}_{\mathrm{o}}-\mathbf{x}_{\mathrm{o}}$, where $\hat{\mathbf{x}}_{\mathrm{o}}$ is the maximum-likelihood estimator of the initial state. Starting from the same initial state we ran the system 5,000 times independently, each dot representing the estimation error of one run. The uncertainty ellipsoid (black) corresponds to the standard deviation of $\tilde{\mathbf{x}}$ in any direction.

Fig. 1c. The corresponding minimum energies are shown in Fig. 1d. The control energy surface for all normalized desired states is an ellipsoid, implying that the required energy varies drastically as we move the system in different directions.

As real systems normally function near a stable state—that is, all eigenvalues of A are negative[37]—the control energy $\mathcal{E}(\tau)$ decays quickly to a nonzero stationary value when the control time τ increases[16]. Henceforth we focus on the control energy $\mathcal{E}\equiv\mathcal{E}(\tau\to\infty)$ and the controllability Gramian $G\equiv G_{\mathrm{c}}(\tau\to\infty)$.

Given a network A and an input matrix B, the controllability Gramian G is unique, embodying all properties related to the control of the system. To uncover the directions of the state space

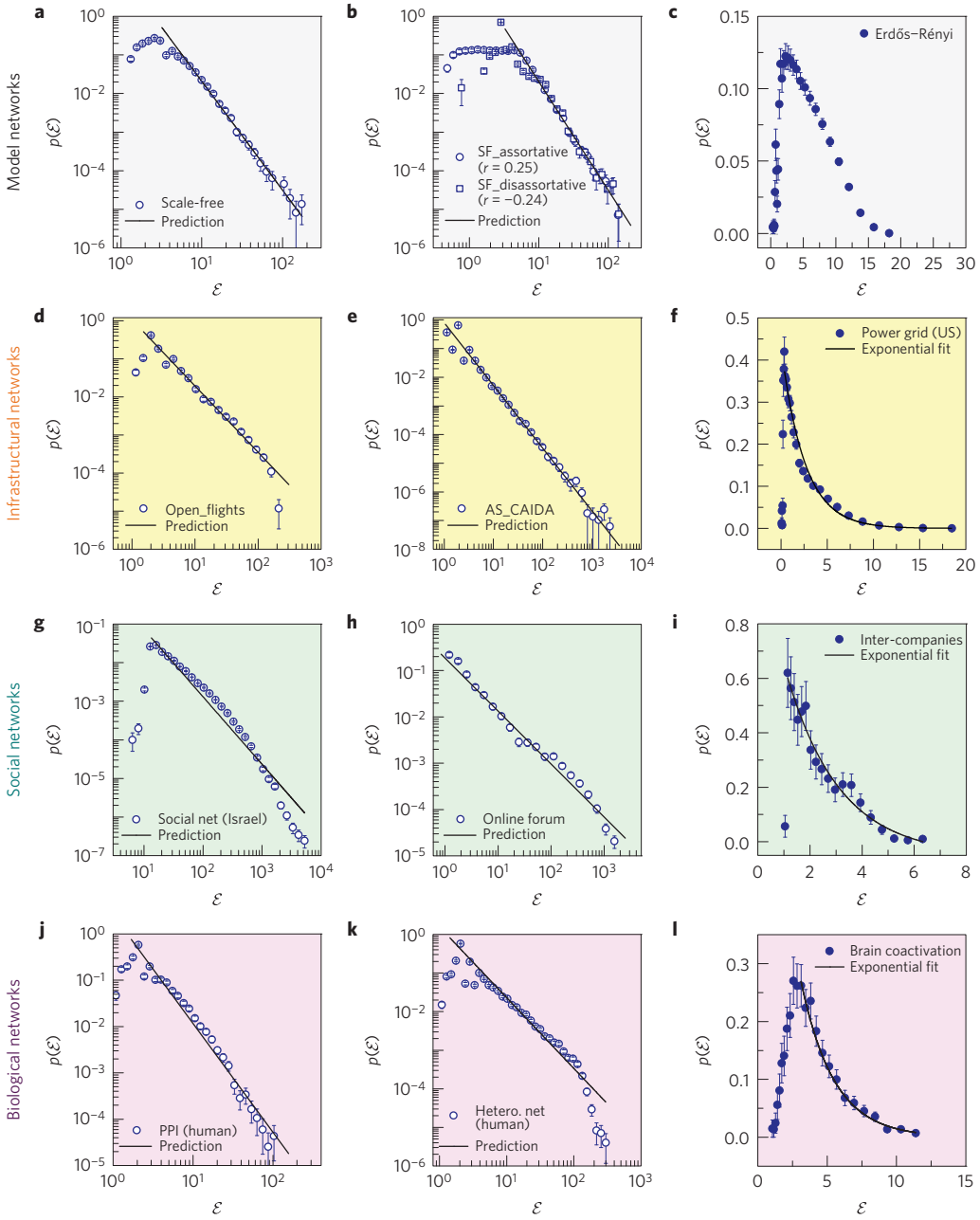

Figure 2 | Controlling a network through all nodes ($N_D = N$). Distribution $p(\mathcal{E})$ of eigen-energies required to control several model and real systems. **a**, Scale-free model network without degree–degree correlation. **b**, Scale-free model networks with positive ($r = 0.25$) or negative ($r = -0.24$) degree–degree correlation. **c**, Erdős–Rényi model network. **d**, Airline transportation network. **e**, Internet AS-level network. **f**, US power grid network. **g**, Israeli social network. **h**, User-interaction network of an online forum. **i**, Interlocking network of Norwegian companies. **j**, Human protein–protein interaction network. **k**, Human heterogeneous network. **l**, Functional coactivation network of the human brain. The straight lines show prediction (4) and the error bars represent standard deviations. For model networks the edges' weights A_{ij} are uniformly drawn from [0,1]. The self-loops $A_{ii} = -\sum_j A_{ij} - \delta$, where $\delta = 0.25$, representing a small perturbation to diagonal entries, to ensure that the network is stable. The data sources and basic characteristics of these networks are discussed in Supplementary Section VI A.

requiring different energies, we explore the eigen-space of G. Denote by $\mathcal{E}_i$ the eigen-energies (that is, the minimum energy required to drive the network to the eigen-directions of G). According to equation (2) $\mathcal{E}_i = 1/\mu_i$, with μ_i corresponding to the eigenvalues of G. Generally, the energy surface for a network with N nodes is a super-ellipsoid spanned by the N eigen-energies of G. To determine the distribution of these eigen-energies we decompose the adjacency matrix as $A = V\Lambda V^{\mathrm{T}}$, where V represents the eigenvectors of A and $\Lambda = \mathrm{diag}\{-\lambda_1, -\lambda_2, \ldots, -\lambda_N\}$ are the eigenvalues. For stable undirected networks all eigenvalues of A are negative, thus we denote the eigenvalues by $-\lambda_i$ so that the absolute eigenvalues are $\lambda_i > 0$ for all i. We sort the absolute eigenvalues in ascending order $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_N$, finding that (Supplementary Section I)

$$G = V[(V^{\mathrm{T}}BB^{\mathrm{T}}V) \circ C]V^{\mathrm{T}} \tag{3}$$

where $\circ$ denotes Hadamard product, defined as $(X \circ Y)_{ij} = X_{ij}Y_{ij}$, and C is a matrix with entries $C_{ij} = 1/(\lambda_i + \lambda_j)$. For a given

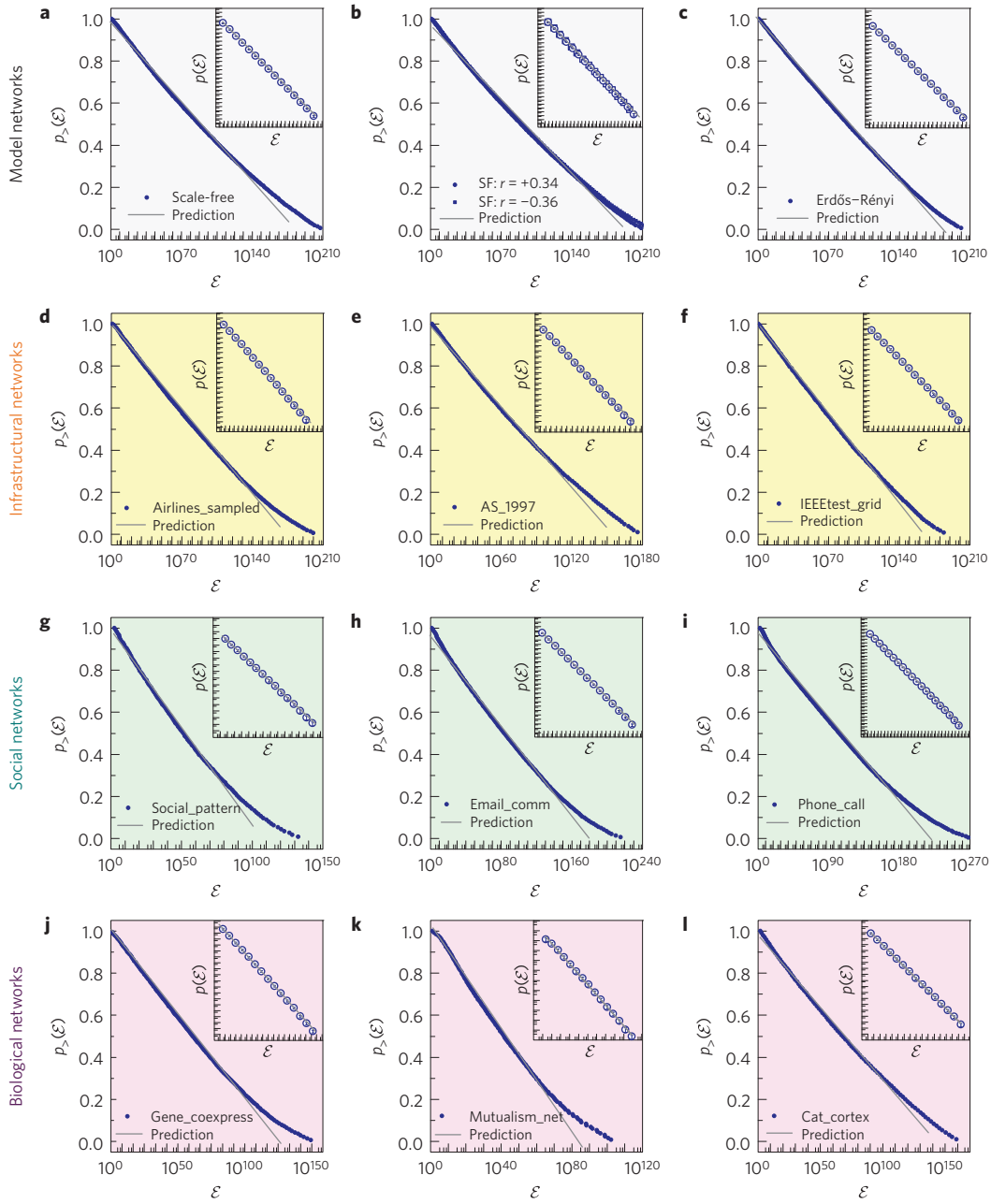

Figure 3 | Controlling a network through a single node ($N_D = 1$). Panels show the complementary cumulative distribution $p_>(\mathcal{E})$ of eigen-energies required to control several model and real systems. **a**, Scale-free model network without degree–degree correlation. **b**, Scale-free model networks with positive ($r = 0.34$) or negative ($r = -0.36$) degree–degree correlation. **c**, Erdős–Rényi model network. **d**, Sampled airline network. **e**, Internet AS-level network in 1997. **f**, IEEE power grid test network. **g**, Human-contact network of the ACM Hypertext 2009 conference. **h**, Email interaction network. **i**, Phone call network between different countries. **j**, Connected component of the human gene-coexpression network. **k**, Mutualism ecological network in Mauritius. **l**, Inter-region network of cat cortex. The insets are log–log plots of probability distributions $p(\mathcal{E})$ with logarithmic binning[3]. The straight lines show prediction (5) and the error bars represent standard deviations. The data sources and basic characteristics of these networks are discussed in Supplementary Section VI A.

network, (3) captures the impact of the input matrix B on the control properties of the system, allowing us to analyse the distribution of eigen-energies for different numbers of driver nodes and determine the required energy for each direction.

Controlling a system through all nodes

If we can control all nodes (that is, $N_D = N$), B becomes a unit diagonal matrix. In this case $G = V\,\mathrm{diag}\{1/2\lambda_i\}V^{\mathrm{T}}$ and the eigen-directions of the controlled system are the same as the network's eigenvectors. Thus $\mathcal{E}_i = 2\lambda_i$ and $p(\mathcal{E}) = (1/2)p(\lambda)$, meaning the distribution of eigen-energies is proportional to the distribution of the network's absolute eigenvalues. We add self-loops as $A_{ii} = -(\delta + \sum_{j=1}^{N} A_{ij})$, where $\delta > 0$ is a small perturbation to ensure that all eigenvalues of A are negative. This scheme has been widely used in previous studies on dynamical processes taking place on networks, such as opinion dynamics[29], synchronization[38] and control[16]. For networks with degree distribution[1–3] $p(k) \sim k^{-\gamma}$ the distribution of the absolute eigenvalues of A also obeys a power

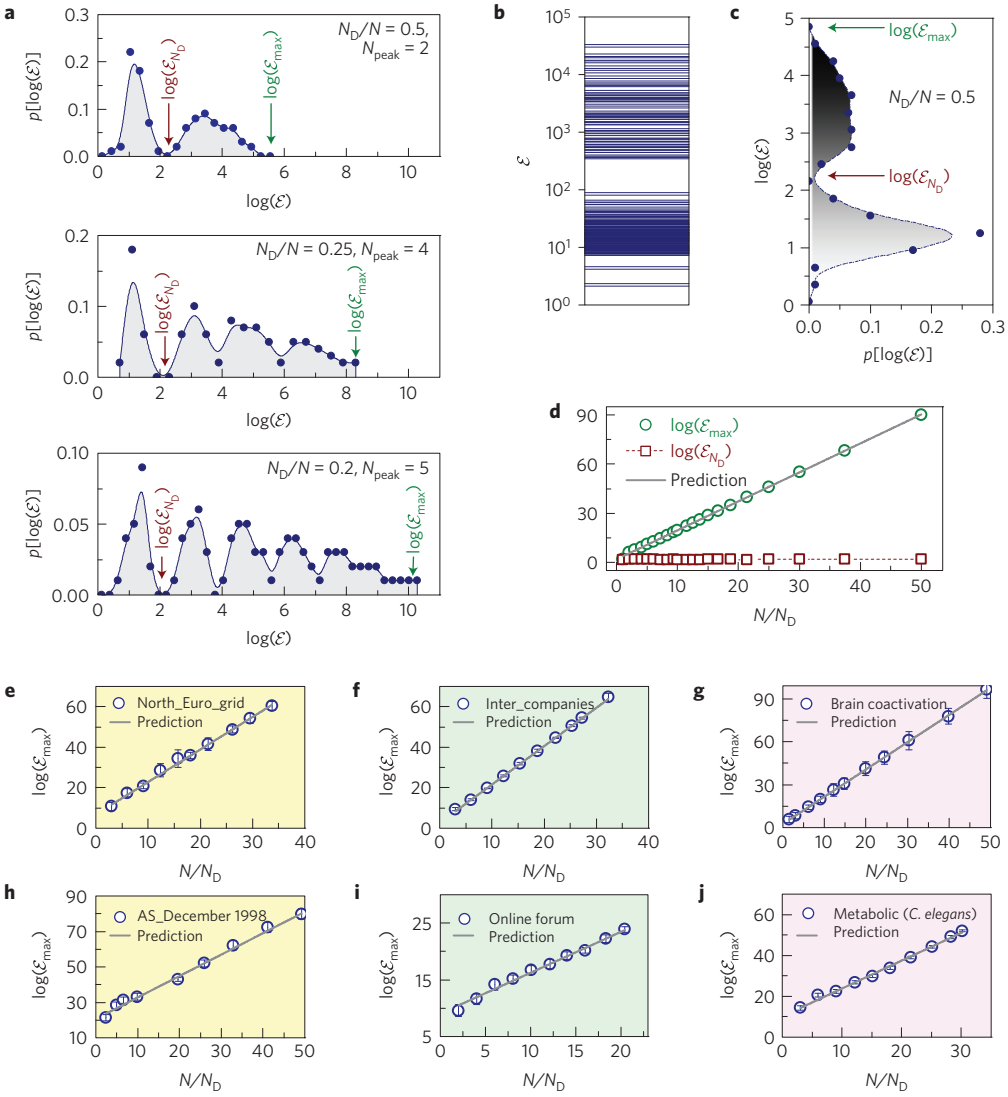

Figure 4 | Controlling a network through a finite fraction of its nodes. a, Multi-peak distributions $p[\log(\mathcal{E})]$ for $1 < N_D < N$, where the dots represent numerical results. The solid curves and shaded areas are smoothed for illustration. $\mathcal{E}_{N_D}$ denotes the boundary of the first energy band that contains N_D eigen-energies. $\mathcal{E}_{max}$ is the maximum control energy corresponding to the most difficult direction. The number of peaks is $N_{peak} = \text{int}[N/N_D]$ (see also Supplementary Fig. 3). **b**,**c**, The eigen-energy spectrum for controlling the network (**b**). There is a gap in the logarithmic scale between the N_Dth and the (N_D+1)th smallest eigen-energies, which leads to the two-peak distribution $p[\log(\mathcal{E})]$ shown in **c**. **d**, $\log(\mathcal{E}_{max})$ and $\log(\mathcal{E}_{N_D})$ as functions of N/N_D, indicating that $\mathcal{E}_{max} \sim e^{N/N_D}$ whereas $\mathcal{E}_{N_D}$ depends weakly on N_D. **e**–**j**, We also test the prediction in real networks: the North European power grid network (**e**), the interlocking network of Norwegian companies (**f**), the functional coactivation network of the human brain (**g**), an Internet AS-level network in December 1998 (**h**), the user-interaction network of an online forum (**i**) and the metabolic network of *Caenorhabditis elegans* (**j**). The straight lines show the prediction $\mathcal{E}_{max} \sim e^{N/N_D}$ and the error bars represent standard deviations. The data sources and basic characteristics of these networks are discussed in Supplementary Section VI A.

law[39,40] $p(\lambda) \sim \lambda^{-\gamma}$ (see Supplementary Section II A). Consequently,

$$p(\mathcal{E}) \sim \mathcal{E}^{-\gamma} \tag{4}$$

indicating that the system can be easily driven in most directions of the state space, requiring a small $\mathcal{E}$. A few directions require considerable energy and the most difficult direction needs[41] $\mathcal{E}_{max} \sim N^{1/(\gamma-1)}$. The fact that $\mathcal{E}_{max}$ is sublinear in N for $\gamma > 2$ indicates that, when $N_D = N$, the energy density $\mathcal{E}/N_D$ remains bounded. In Fig. 2 we test the prediction (4) for several model, infrastructural, social and biological networks. We find that $p(\mathcal{E})$ follows a power law for uncorrelated or correlated scale-free model networks (Fig. 2a,b), the airline transportation network (Fig. 2d), the Internet AS-level network (Fig. 2e), an Israeli social network (Fig. 2g), the user-interaction network of an online forum (Fig. 2h), the human protein–protein interaction network (Fig. 2j) and the human heterogeneous network (Fig. 2l), in line with the prediction (4). In contrast, for several networks with bounded degree distribution, such as the Erdős–Rényi random network (Fig. 2c), the US power grid network (Fig. 2f), the interlocking network of Norwegian companies (Fig. 2i) and the functional coactivation network of the human brain (Fig. 2l), $p(\mathcal{E})$ is also bounded, as predicted by $\gamma \to \infty$ in (4). Such networks require even less energy to control their progress in their most difficult direction. Taken together, we find that for $N_D = N$ the distribution of eigen-energies is uniquely determined by network topology, and we lack significant energetic barriers for control.

Controlling a system through a single node

If there is not degeneracy in A's eigenvalues we can control an undirected network by driving only a single node[11,42]. In this case $V^{\mathrm{T}}BB^{\mathrm{T}}V = \{V_{ih}V_{jh}\} \sim \mathcal{O}(1/N)$, where h is the index of the chosen driver node. Thus $V^{\mathrm{T}}BB^{\mathrm{T}}V$ can be viewed as a small perturbation to the matrix C in (3). The statistical behaviour of the eigenvalues of G is mainly determined by the eigenvalues of C, which can be approximated by Cholesky factors[43]. As mentioned above, for networks with $p(k) \sim k^{-\gamma}$, the distribution of the absolute eigenvalues of A also follows a power law, providing the ith eigenvalue $\lambda_i \sim (N/(N+1-i))^{1/(\gamma-1)}$ (Supplementary Section II B). If $\gamma \to 0$—that is, for extremely heterogeneous networks[15,44]—the eigenvalue gaps $g_i \equiv \lambda_{i+1} - \lambda_i$ are identical. For $\gamma \to \infty$ (homogeneous networks), g_i is again uniform. Hence it is reasonable to assume $g_i = g$ for all i, allowing us to analytically obtain the distribution of eigen-energies as $p(\mathcal{E}) \sim 1/(1+1/\mathcal{E})\mathcal{E}^{-1}$ (see Supplementary Section III A, B). Therefore,

$$p(\mathcal{E}) \sim \mathcal{E}^{-1} \tag{5}$$

for large $\mathcal{E}$. Equation (5) predicts that, to drive a stable network of N nodes with a single driver node, the most difficult direction in the state space requires $\mathcal{E}_{\max} \sim \mathrm{e}^N$ energy (Supplementary Section III C). This exponential N-dependence makes the control of large networks in the most difficult direction energetically infeasible. For validation we also consider the complementary cumulative distribution $p_>(\mathcal{E}) = \int_{\mathcal{E}}^{\mathcal{E}_{\max}} p(\mathcal{E}')\mathrm{d}\mathcal{E}'$. Based on (5) we obtain $p_>(\mathcal{E}) \sim (\ln \mathcal{E}_{\max} - \ln \mathcal{E})$, decreasing linearly with $\ln \mathcal{E}$. We test our prediction on several network models (Fig. 3a–c) and real networks (Fig. 3d–l), finding that the corresponding eigen-energies span over a hundred orders of magnitude and this exceptional range of variations is reasonably well approximated by (5) for both $p_>(\mathcal{E})$ and $p(\mathcal{E})$. Taken together, if we attempt to control a network from a single node ($N_{\mathrm{D}} = 1$), the required energy varies enormously for different directions, almost independently of the network structure, making some directions prohibitively expensive energetically.

Controlling a system through a finite fraction of its nodes

When $p(\mathcal{E}) \sim \mathcal{E}^{-\gamma}$, the distribution $p(\hat{\mathcal{E}}) \sim \mathrm{e}^{(1-\gamma)\hat{\mathcal{E}}}$, where $\hat{\mathcal{E}} \equiv \ln \mathcal{E}$. Thus, if $N_{\mathrm{D}} = N$, $p(\hat{\mathcal{E}})$ is an exponential (one-peak) distribution for $\gamma > 2$ in (4); if $N_{\mathrm{D}} = 1$, as $p(\mathcal{E}) \sim \mathcal{E}^{-1}$ in (5), $p(\hat{\mathcal{E}})$ is a uniform distribution. To understand the transition from (4) for $N_{\mathrm{D}} = N$ to (5) for $N_{\mathrm{D}} = 1$, we investigate the distribution $p(\hat{\mathcal{E}})$ when $1 < N_{\mathrm{D}} < N$ (that is, when we try to control a system through a finite fraction of its nodes). In this case we find that $p(\hat{\mathcal{E}})$ has multiple peaks (Fig. 4a), which are induced by gaps in the eigen-energy spectrum (Fig. 4b). For $N_{\mathrm{D}}/N = 0.6$, a gap separates the eigen-energies into two bands, such that the lower band contains N_{D} eigen-energies. This gap leads to two peaks in the distribution $p(\hat{\mathcal{E}})$, as shown in Fig. 4c. When we have fewer driver nodes (N_{D}/N decreases), the number of peaks N_{peak} increases (Fig. 4a). We find that $N_{\mathrm{peak}} = \mathrm{int}[N/N_{\mathrm{D}}]$, predicting $N_{\mathrm{peak}} = 2, 4, 5$ for $N_{\mathrm{D}}/N = 0.5, 0.25, 0.2$, respectively (see also Supplementary Fig. 3). The multi-peak nature of $p(\hat{\mathcal{E}})$ has two important implications. First, the boundary of the first energy band $\mathcal{E}_{N_{\mathrm{D}}}$ varies only weakly with N_{D} (Fig. 4d), indicating that the energy required to move the network within the subspace spanned by the first N_{D} eigen-directions is relatively small. Second, $\hat{\mathcal{E}}$ (that is, $\log \mathcal{E}$) grows linearly from one band to the next (Supplementary Fig. 4). Thus, $\log \mathcal{E}_{\max}$ (the boundary of the last band) is linearly dependent on the number of peaks (that is, $\mathcal{E}_{\max} \sim \mathrm{e}^{N/N_{\mathrm{D}}}$) (Fig. 4d). Controlling a single node induces N peaks in $p(\hat{\mathcal{E}})$, consequently the distribution $p(\hat{\mathcal{E}})$ becomes uniform (Supplementary Fig. 3), resulting in $p(\mathcal{E}) \sim \mathcal{E}^{-1}$ of (5) and $\mathcal{E}_{\max} \sim \mathrm{e}^N$. We numerically test the prediction $\mathcal{E}_{\max} \sim \mathrm{e}^{N/N_{\mathrm{D}}}$ for several real networks (Fig. 4e–j), the result being in excellent agreement with our prediction.

Table 1 | Controlling complex networks with different numbers of driver nodes.

Number of driver nodes	Distribution of eigen-energies	Maximum control energy
$N_{\mathrm{D}} = N$	$p(\mathcal{E}) \sim \mathcal{E}^{-\gamma}$	$\mathcal{E}_{\max} \sim N^{1/(\gamma-1)}$
$1 < N_{\mathrm{D}} < N$	$N_{\mathrm{peak}} = \mathrm{int}[N/N_{\mathrm{D}}]$ for $p[\log(\mathcal{E})]$	$\mathcal{E}_{\max} \sim \mathrm{e}^{N/N_{\mathrm{D}}}$
$N_{\mathrm{D}} = 1$	$p(\mathcal{E}) \sim \mathcal{E}^{-1}$	$\mathcal{E}_{\max} \sim \mathrm{e}^N$

N is the total number of nodes and N_{D} is the number of driver nodes. γ is the exponent of the degree distributions $p(k) \sim k^{-\gamma}$. For large γ the network becomes degree-homogeneous, behaving similarly to a random network.

In Table 1 we summarize our findings about the distribution of eigen-energies and the maximum energy required to control a system towards the most difficult direction.

Implications to observational uncertainty

The results obtained above have direct implications for observability as well. Indeed, consider a system governed by the dynamics

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t)$$

$$\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{w}(t)$$

with an initial state $\mathbf{x}_{\mathrm{o}} \neq 0$, where C is the output matrix and $\mathbf{y}(t)$ are the output signals including measurement noise $\mathbf{w}(t)$, which we assume to be a Gaussian white noise with zero mean and variance one. We aim to estimate $\hat{\mathbf{x}}_o$ of the initial state $\mathbf{x}_{\mathrm{o}}$ while minimizing the difference $\int_0^\tau \|\mathbf{y}(t) - \hat{\mathbf{y}}(t)\|^2 \mathrm{d}t$ between the output $\mathbf{y}(t)$ that is actually observed and the output $\hat{\mathbf{y}}(t) = Ce^{At}\hat{\mathbf{x}}_{\mathrm{o}}$ that would be observed in the absence of noise. With the maximum-likelihood approximation[45], the expectation $\langle \hat{\mathbf{x}}_{\mathrm{o}} \rangle = \mathbf{x}_{\mathrm{o}}$ and the covariance matrix[45] $\langle \tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{T}} \rangle = G_{\mathrm{o}}^{-1}(\tau)$, where $\tilde{\mathbf{x}} \equiv \hat{\mathbf{x}}_{\mathrm{o}} - \mathbf{x}_{\mathrm{o}}$ is the estimation error and $G_{\mathrm{o}}(\tau) = \int_0^\tau \mathrm{e}^{A^{\mathrm{T}}t} C^{\mathrm{T}} C \mathrm{e}^{At} \mathrm{d}t$ is the observability Gramian. Therefore, the variance σ^2 of the approximation in direction $\tilde{\mathbf{x}}$ is

$$\sigma^2(\tau) = \tilde{\mathbf{x}}^{\mathrm{T}} G_{\mathrm{o}}^{-1}(\tau) \tilde{\mathbf{x}} \tag{6}$$

indicating that the estimation uncertainty varies with the direction of the state space. To illustrate this, consider the network in Fig. 1e that moves along the trajectory of Fig. 1g, while we measure the state of the sensor node and plot the noisy output $\mathbf{y}(t)$ in Fig. 1f. With the maximum-likelihood approximation we reconstruct $\mathbf{x}_{\mathrm{o}}$ from $\mathbf{y}(t)$ and show the estimation error $\tilde{\mathbf{x}} \equiv \hat{\mathbf{x}}_{\mathrm{o}} - \mathbf{x}_{\mathrm{o}}$ for thousands of independent runs (Fig. 1h). The estimation variance is different for various directions, forming an uncertainty ellipsoid. Thanks to the duality between $G_{\mathrm{c}}(\tau)$ and $G_{\mathrm{o}}(\tau)$, the control energy for a direction in Fig. 1d represents the estimation variance for the same direction in Fig. 1h.

To be specific, owing to the duality of the controllability Gramian G_{c} in (2) and the observability Gramian G_{o} in (6), we have $\sigma^2 = \mathcal{E}$ for the same direction, implying that the least controllable direction (that is, the direction requiring the most energy) is also least observable (having highest uncertainty). Therefore, our findings about the distribution of eigen-energies apply directly to the distribution of σ^2 along the eigen-directions: if all nodes are sensor nodes ($N_{\mathrm{S}} = N$) we have $p(\sigma^2) \sim (\sigma^2)^{-\gamma}$; if we attempt to observe the system from a single node ($N_{\mathrm{S}} = 1$) we have $p(\sigma^2) \sim (\sigma^2)^{-1}$; and for a finite fraction of sensor nodes ($1 < N_{\mathrm{S}} < N$) the largest observational uncertainty $\sigma^2_{\max}$ decreases exponentially when the number of sensor nodes increases (that is, $\sigma^2_{\max} \sim \mathrm{e}^{N/N_{\mathrm{S}}}$).

Beyond the degree distribution

Real networks have a number of additional properties that are not encoded by their degree distributions, such as local clustering[46],

degree correlations[47] and community structure[48] (Supplementary Table 1). To assess the impact of these topological characteristics we perform the degree-preserved randomization[49] on each network, eliminating local clustering, degree correlations and modularity. We find that the distribution of eigen-energies required to drive each randomized network follows the predictions (4) and (5) (Supplementary Figs 6 and 7), indicating that degree distribution is the main factor determining $p(\mathcal{E})$. When the number of driver nodes increases, the maximum control energy for the randomized networks decreases exponentially, as predicted earlier (Supplementary Fig. 8). We also validate the predictions (4) and (5) on model networks with positive or negative degree correlations (Figs 1b and 2b). All the tests indicate that the strength of local clustering, degree correlations or community structure have only minor influence on the behaviour of control energy. Consequently, our calculations for uncorrelated networks capture the correct fundamental dependence of control energy for real networks.

Many real networks have dead ends (that is, nodes with one degree), which can undermine the stability of complex systems[50]. To test the impact of dead ends on control energy we explored several real networks that contain a considerable number of one-degree nodes (see Supplementary Table 1). As shown in Figs 2–4 and Supplementary Figs 6–8, the predictions are robust against such dead ends.

Discussion

The energy required for control is a significant issue for practical control of complex systems. By exploring the eigen-space of controlled systems we found that if all nodes of a system are directly driven, the eigen-energies can be heterogeneous or homogeneous, depending on the structure of underlying networks. Yet, if we wish to control a system through a single node, the eigen-energies are enormously heterogeneous, almost independently of the network structure. Finally, if a finite fraction of nodes are driven, the maximum control energy decays exponentially with the increasing number of driver nodes. Taken together, our results indicate that even if controllable, most systems still have directions which are energetically inaccessible, suggesting a natural mechanism to avoid undesirable states. Indeed, many complex systems, such as transcriptional networks for gene expression[51] and sensorimotor systems for motion control[52], need only to function in a low-dimensional subspace. Owing to the duality between controllability and observability, our results also imply that, if we monitor only a small fraction of nodes, the observation can be extremely unreliable in certain directions of the phase space.

It is worth noting that linear dynamics captures the behaviour of nonlinear systems in the vicinity of their equilibria. The formalism (1) has been widely used to model diverse networked systems[14,18,24,29,36] (see also Supplementary Section VII A, B), allowing us to reveal the role of the network topology on the fundamental control properties of complex systems[10,11,14–24]. Indeed, if the linearized system (1) is controllable, the original nonlinear system is locally controllable[53]. The corresponding control energy is also highly heterogeneous for different directions, if we constrain the system's trajectory to be local (Supplementary Section VII C). Moreover, if the linearized dynamics of a nonlinear system is controllable along a specific trajectory, the original nonlinear system is locally controllable along the same trajectory[53]. This implies that our results can be potentially extended to describe control properties of nonlinear systems in the vicinity of their stability basin[50,54]. Yet, in this case, the linearized dynamics becomes time-varying, and the required energy for controlling time-varying systems remains an open problem that deserves future attention.

Received 17 December 2014; accepted 30 June 2015;
published online 10 August 2015

References

1. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74,** 47–97 (2002).
2. Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function* (Cambridge Univ. Press, 2010).
3. Newman, M. E. J. *Networks: An Introduction* (Oxford Univ. Press, 2010).
4. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. Complex networks: Structure and dynamics. *Phys. Rep.* **424,** 175–308 (2006).
5. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, 2008).
6. Barzel, B. & Barabási, A.-L. Universality in network dynamics. *Nature Phys.* **9,** 673–681 (2013).
7. Rugh, W. J. *Linear System Theory* (Prentice-Hall, 1996).
8. Sontag, E. D. *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer, 1998).
9. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
10. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473,** 167–173 (2011).
11. Yuan, Z., Zhao, C., Di, Z., Wang, W.-X. & Lai, Y.-C. Exact controllability of complex networks. *Nature Commun.* **4,** 2447 (2013).
12. Sorrentino, F., di Bernardo, M., Garofalo, F. & Chen, G. Controllability of complex networks via pinning. *Phys. Rev. E* **75,** 046103 (2007).
13. Yu, W., Chen, G. & Lü, J. On pinning synchronization of complex dynamical networks. *Automatica* **45,** 429–435 (2009).
14. Rajapakse, I., Groudine, M. & Mesbahi, M. Dynamics and control of state-dependent networks for probing genomic organization. *Proc. Natl Acad. Sci. USA* **108,** 17257–17262 (2011).
15. Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nature Phys.* **8,** 568–573 (2012).
16. Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H. & Li, B. Controlling complex networks: How much energy is needed? *Phys. Rev. Lett.* **108,** 218703 (2012).
17. Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. *Phys. Rev. Lett.* **110,** 208701 (2013).
18. Pasqualetti, F., Zampieri, S. & Bullo, F. Controllability metrics, limitations and algorithms for complex networks. *IEEE Trans. Control Netw. Syst.* **1,** 40–52 (2014).
19. Tang, Y., Gao, H., Zou, W. & Kurths, J. Identifying controlling nodes in neuronal networks in different scales. *PLoS ONE* **7,** e41375 (2012).
20. Jia, T. *et al.* Emergence of bimodality in controlling complex networks. *Nature Commun.* **4,** 2002 (2013).
21. Ruths, J. & Ruths, D. Control profiles of complex networks. *Science* **343,** 1373–1376 (2014).
22. Menichetti, G., Dall'Asta, L. & Bianconi, G. Network controllability is determined by the density of low in-degree and out-degree nodes. *Phys. Rev. Lett.* **113,** 078701 (2014).
23. Summers, T. H., Cortesi, F. L. & Lygeros, J. On submodularity and controllability in complex dynamical networks. Preprint at http://arXiv.org/abs/1404.7665v2 (2014).
24. Tzoumas, V., Rahimian, M. A., Pappas, G. J. & Jadbabaie, A. Minimal actuator placement with optimal control constraints. Preprint at http://arXiv.org/abs/1503.04693 (2015).
25. Cornelius, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nature Commun.* **4,** 1942 (2013).
26. Whalen, A. J., Brennan, S. N., Sauer, T. D. & Schiff, S. J. Observability and controllability of nonlinear networks: The role of symmetry. *Phys. Rev. X* **5,** 011005 (2015).
27. Menolascina, F. *et al. In-vivo* real-time control of protein expression from endogenous and synthetic gene networks. *PLoS Comput. Biol.* **10,** e1003625 (2014).
28. Rahmani, A., Ji, M., Mesbahi, M. & Egerstedt, M. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Control Optim.* **48,** 162–186 (2009).
29. Acemoglu, D., Ozdaglar, A. & ParandehGheibi, A. Spread of (mis)information in social networks. *Games Econ. Behav.* **70,** 194–227 (2010).
30. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Natl Acad. Sci. USA* **110,** 2460–2465 (2013).
31. Yang, Y., Wang, J. & Motter, A. E. Network observability transitions. *Phys. Rev. Lett.* **109,** 258701 (2012).
32. Pinto, P. C., Thiran, P. & Vetterli, M. Locating the source of diffusion in large-scale networks. *Phys. Rev. Lett.* **109,** 068702 (2012).
33. Scheffer, M. *et al.* Anticipating critical transitions. *Science* **338,** 344–348 (2012).
34. Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303,** 799–805 (2004).
35. Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N. & Barabási, A.-L. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427,** 839–843 (2004).

36. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81,** 591–646 (2009).
37. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1974).
38. Pecora, L. M. & Carroll, T. L. Master stability functions for synchronized coupled systems. *Phys. Rev. Lett.* **80,** 2109–2112 (1998).
39. Chung, F., Lu, L. & Vu, V. Spectra of random graphs with given expected degrees. *Proc. Natl Acad. Sci. USA* **100,** 6313–6318 (2003).
40. Kim, D. & Kahng, B. Spectral densities of scale-free networks. *Chaos* **17,** 026115 (2007).
41. Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85,** 4626–4628 (2000).
42. Cowan, N. J., Chastain, E. J., Vilhena, D. A., Freudenberg, J. S. & Bergstrom, C. T. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLoS ONE* **7,** e38398 (2012).
43. Antoulas, A. *Approximation of Large-Scale Dynamical Systems* (SIAM, 2009).
44. Del Genio, C., Gross, T. & Bassler, K. All scale-free networks are sparse. *Phys. Rev. Lett.* **107,** 178701 (2011).
45. Kailath, T., Sayed, A. & Hassibi, B. *Linear Estimation* (Prentice-Hall, 2000).
46. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (1998).
47. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89,** 208701 (2002).
48. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99,** 7821–7826 (2002).
49. Xulvi-Brunet, R. & Sokolov, I. M. Reshuffling scale-free networks: From random to assortative. *Phys. Rev. E* **70,** 066102 (2004).
50. Menck, P. J., Heitzig, J., Kurths, J. & Schellnhuber, H. J. How dead ends undermine power grid stability. *Nature Commun.* **5,** 3969 (2014).
51. Müller, F.-J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478,** E4 (2011).
52. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5,** 1226–1235 (2002).
53. Coron, J.-M. *Control and Nonlinearity* (American Mathematical Society, 2009).
54. Menck, P. J., Heitzig, J., Marwan, N. & Kurths, J. How basin stability complements the linear-stability paradigm. *Nature Phys.* **9,** 89–92 (2013).

Acknowledgements

We thank E. Guney, C. Song, J. Gao, M. T. Angulo, S. P. Cornelius, B. Coutinho and A. Li for discussions. This work was supported by Army Research Laboratories (ARL) Network Science (NS) Collaborative Technology Alliance (CTA) grant ARL NS-CTA W911NF-09-2-0053; DARPA Social Media in Strategic Communications project under agreement number W911NF-12-C-002; the John Templeton Foundation: Mathematical and Physical Sciences grant number PFI-777; European Commission grant numbers FP7 317532 (MULTIPLEX) and 641191 (CIMPLEX).

Author contributions

All authors designed and performed the research. G.Y. and G.T. carried out the numerical calculations. G.Y. did the analytical calculations and analysed the empirical data. G.T., B.B., J.-J.S., Y.-Y.L. and A.-L.B. analysed the results. G.Y. and A.-L.B. were the main writers of the manuscript. G.T., B.B. and Y.-Y.L. edited the manuscript. G.Y. and G.T. contributed equally to this work.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.-L.B.

Competing financial interests

The authors declare no competing financial interests.